AGENT BASED DOCUMENT ANALYSIS, Dale J. Cox, Adam J. Kennedy, Jeremiah S. Parisien, Munehiro Kura, Jack R. Hagemeister*, Washington State University, School of Electrical Engineering and Computer Science, Pullman, WA 99163, jackrh@eecs.wsu.edu

Classifying large data sets of text can result in algorithms that require extensive processing power, especially as the data set grows. In a world that has grown increasingly reliant upon electronic means of communication and collaboration, the amount of information available for classification and analysis is increasingly vast. As such, alternate algorithms for categorization of data may be required in order to analyze the information in an efficient manner. In addition to simply classifying data, prioritization of this information becomes vital to allow the documents to be useful to a user.

Agent based categorization presents an alternate paradigm to the traditional algorithms. Our work, conducted in collaboration with researchers from the National Visualization and Analytics Center (NVAC) at Pacific Northwest National Laboratory, uses the JADE agent framework to create an agent-based tool for classifying and prioritizing documents based on user preferences.

The software we have created employs various types of agents to identify documents of importance, and reports this information to the user. The design of the system supports continuous data updates, therefore there never exists a "finished" solution; instead the system continuously monitors and acts upon a stream of documents in an effort to find documents of value to the user. The use of agent-based techniques allows us to employ multiple strategies for both classification and prioritization, and allows different strategies to dominate as the content of the new stream varies with time.

We will discuss the overall agent framework (built using the JADE agent framework to allow for the creation of agents that are inherently distributed); the implementation of simple document classification and prioritization agents; and how these simple agents can be expanded to create a rich text processing "ecosystem."   We will also discuss our approach to using competition among agents to be associated with individual documents, and to "bid" for the right to bring the document (or collection of documents) to the user's attention.

This system, with slight modification, has many potential applications. These include security, intelligence gathering, as well as monitoring chat rooms. Furthermore, there exists the possibility for personal applications, such as filtering email or observing trends on social networking sites, such as Facebook. Due to its continuous nature, the system can be used in situations where constant monitoring is ideal, and can be deployed on large scales.